Stat Methods Appl (2015) 24:305–306
DOI 10.1007/s10260-015-0300-4

DISCUSSION

P. Secchi, S. Vantini and V. Vitelli: Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of Milan

Piotr Kokoszka

Accepted: 21 February 2015 / Published online: 28 March 2015

I thank the authors and the editorial board for inviting me to comment on this work. I would like to congratulate the authors on bringing in a new tool set for the analysis of functional data indexed by spatial locations. Data of this type have been attracting attention of the FDA research community for some time, but the focus has been of functional principal component (FPC) expansions to form a suitable temporal basis. My comments will focus on providing several relevant references to my own research, which are related to this work. Many other researchers have contributed to this field, but I do not aim at a comprehensive review.

Delicado et al. (2010) provide a very good review. They review research on geostatistical functional data (focusing on kriging), point processes associated with functional data, and functional areal data. My own work in this area has focused on geostatistical functional data, which can be denoted as $X(\mathbf{s}_k; t)$, where $\mathbf{s}_k$ is the spatial location and t is time. Using this notation, the Erlang data can be written as $E(\mathbf{x}_k; t)$, where the $\mathbf{x}_k$ are the locations with almost complete records. Gromenko et al. (2012) study the estimation of the mean function μ and the FPC's v_j in the Karhunen–Loéve representation

$$X(\mathbf{s}_k; t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j(\mathbf{s}_k) v_j(t)$$

which is analogous to the representation (3) which the authors use. As noted above, the difference is that the v_j are the FPC's rather that the wavelet-like functions ψ_j. In Gromenko et al. (2014b), this work is extended to handle functions which contain

P. Kokoszka (✉)
Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA
e-mail: Piotr.Kokoszka@colostate.edu
URL: http://www.stat.colostate.edu/piotr/

large gaps in the temporal coverage. Due to the dense spatial coverage of the Erlang data, the locations with incomplete functions can just be removed, but it might be useful to extend the treelet method to deal with situations where spatial coverage is sparse and ignoring many locations would lead to a substantial loss of information. It should be noted that if the locations $\mathbf{s}_k$ are distributed with large spatial gaps, then the estimated FPCs are in general not consistent estimators of the population FPC's v_j, Hörmann and Kokoszka (2013). This work is also reported in Chapters 17 and 18 of Horváth and Kokoszka (2012). A theoretical investigation of the treelet method in this direction might be of interest.

The present work covers only 2 weeks, it would be interesting to see how the results would change if a full year of data were used. Clearly, local events, like a Saturday afternoon fair, would not show up in the analysis, but the wavelet power spectrum might be concentrated on fewer shapes, the spectrum in Fig. 6 shows about ten treelet shapes that are almost as significant as the Saturday fair shape. If longer time coverage is to be considered in future research, it might be useful to consider the model of the form $E_n(\mathbf{x}; t)$, where n is the week index. Such a model might help study an annual pattern, which might be useful in urban planning. An approach of this type, applied to a different problem, is used by Gromenko et al. (2014a).

I think many FDA problems that have been studied by means of functional principal component analysis could benefit from a fresh analysis based on treelets or similar custom build basis. For example, Kokoszka and Reimherr (2013) consider the shapes of cumulative intraday returns on US stocks by decomposing them individually into the FPC's. It would be interesting to see the results of their analysis if the stocks are treated as a panel ($\mathbf{x}$ would denote corporations arranged according to some criterion), and what kind of natural intraday shapes ψ_k a treelet or a similar analysis would yield.

References

Delicado P, Giraldo R, Comas C, Mateu J (2010) Statistics for spatial functional data: some recent contributions. Environmetrics 21:224–239

Gromenko O, Kokoszka P, Zhu L, Sojka J (2012) Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. Ann Appl Stat 6:669–696

Gromenko O, Kokoszka P, Reimherr M (2014a) Detection of change in the mean function over a spatial region. Technical report, Colorado State University, Fort Collins

Gromenko O, Kokoszka P, Sojka J (2014b) Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. Technical report, Colorado State University, Fort Collins

Hörmann S, Kokoszka P (2013) Consistency of the mean and the principal components of spatially distributed functional data. Bernoulli 19:1535–1558

Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York

Kokoszka P, Reimherr M (2013) Predictability of shapes of intraday price curves. Econom J 16:285–305